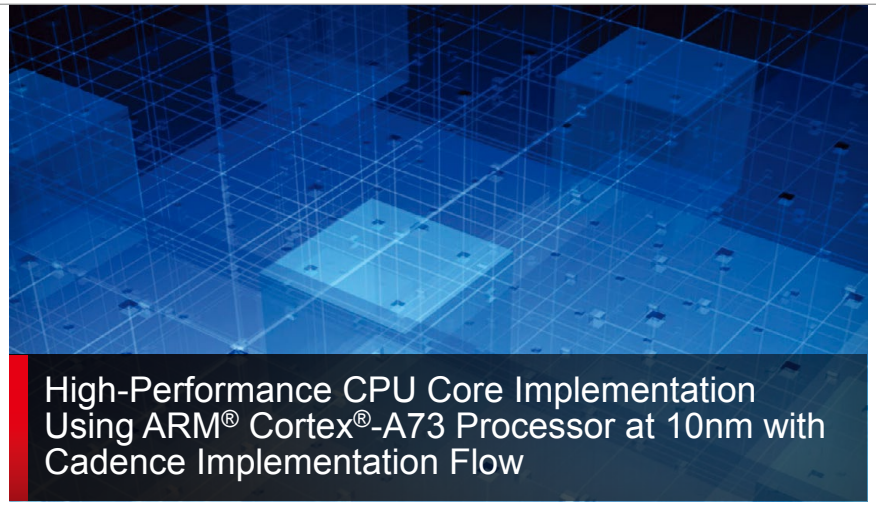
High-Performance CPU Core Implementation Using ARM® Cortex®-A73 Processor at 10nm with Cadence Implementation Flow

ARM / Cadence

ABSTRACT

This will be joint presentation between Cadence and ARM based on a recent 10nm tapeout project implemented with Cadence EDA tools and ARM IP. This presentation will aim to educate attendees on the challenges and implementation strategies for going through different parts of the implementation flow - Synthesis, Floorplannning, Power planning, Detailed Place and Route and Timing/Power Signoff . It will provide Power grid, CLocking and EM awareness strategies and design tips related to 10nm FinFET structures tips and techniques for optimizing implementation of ARM IP on 10nm process. Key ARM and Cadence engineers that have gone through the challenges of implementing a 10nm CPU and will share this hands on experience during the presentation.



High-Performance CPU Core Implementation Using ARM® Cortex®-A73 Processor at 10nm with Cadence Implementation Flow

Shawn Hung – Principal Design Engineer, ARM CPU group
Paddy Mamtora – Group Director, Product Engineering, Cadence
TSMC OIP 2016
San Jose, CA
22 September 2016

cadence

Overview

- ARM® + Cadence Collaboration
- Introduction to Cortex®-A73 Processor Core
- 10nm Design Challenges
- 10nm Design Flow Exploration for High-End ARM CPU Implementation
- Summary and Further Work
- Q&A

cadence

ARM and Cadence
Collaborating on 10nm

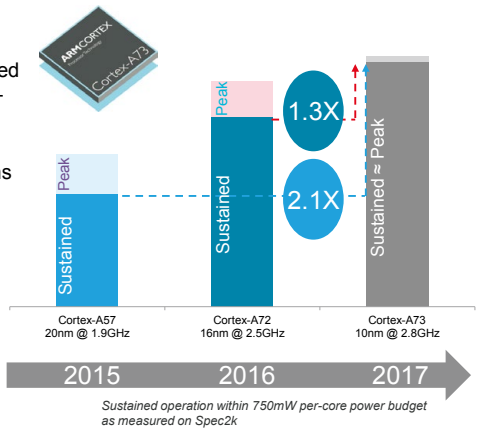
- A collaborative program involving ARM, TSMC, and Cadence to develop the 10nm process ecosystem
- Ensure ARM next-generation CPU IP such as Cortex-A73 is optimized for advanced process
- Ensure EDA ecosystem and design flows are ready for lead partners
- Provide feedback on power, performance, and area bottlenecks to ARM design teams
- Identify additional physical IP requirements for optimal PPA
- Educate partners and accelerate next-generation process adoption

Proven successful approach following on from 16nm collaboration

cadence

ARM® Cortex®-A73 CPU
Most efficient ARM premium core maximizes performance

- Highest performance in mobile power envelope
  - High-end performance with reduced area and power delivering best-in-class efficiency
  - Fits premium and mid-range process for cost-sensitive solutions
- New levels of sustained usage for the best user experience
  - Up to 30% increased power efficiency over Cortex-A72
- The smallest premium ARMv8-A processor ever
  - Over 20% smaller than Cortex-A72 in cost-efficient 28nm process



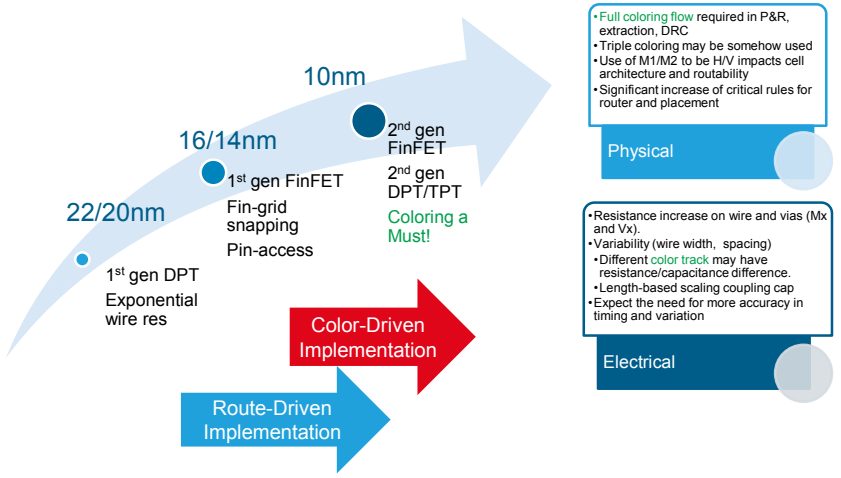
Sustained operation within 750mW per-core power budget as measured on Spec2k

cadence

10nm Design Challenges

cadence

Design Challenges at 10nm



cadence

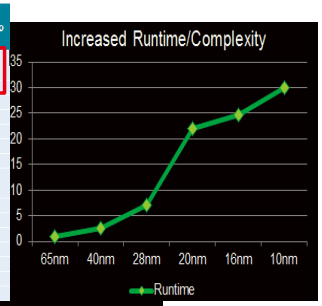
ARM® Cortex® Cores and 10nm Challenges

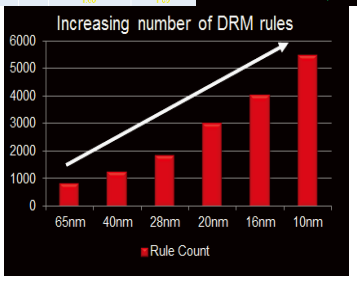
An evolution from 16nm FinFET

- Focus on efficiency
  - Optimize performance within established mobile device power budgets
- Power grid integrity is crucial
  - Meeting electromigration and IR drop targets while maintaining placement density
- Interconnect (wire) resistance continues to dominate
  - Use of optimized physical libraries more critical
- Maximize the benefits of process scaling

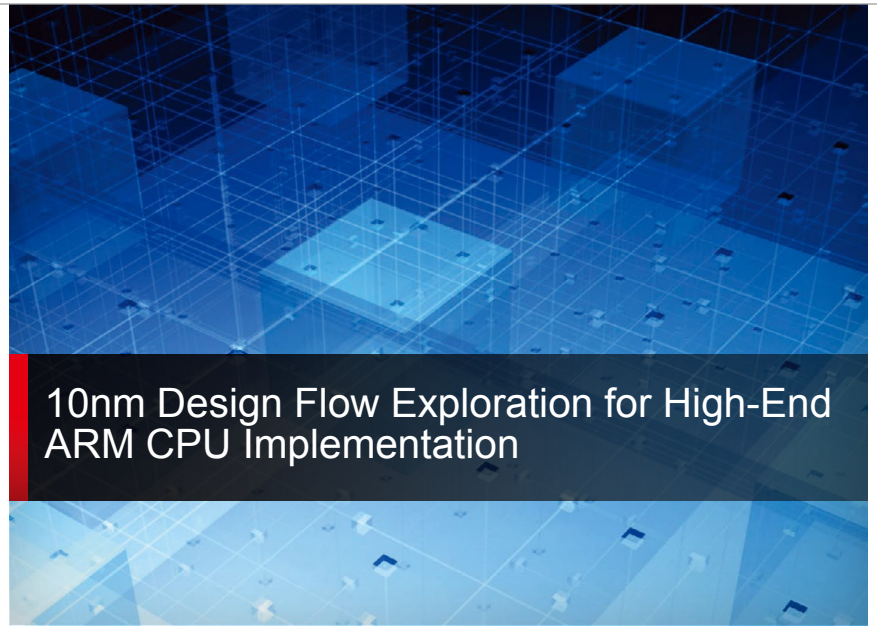
Over 2.5x R↑ Under halved DC Imax ↓

Metal Wire/Via N16	Metal Wire/Via N10	Layer	N10/N16 R Ratio @ Typical/85c	N10/N16 DC I_{max} Ratio @ 100c
		M1	2.95	0.44
Mx/Mxa		M2	2.49	0.48 0.44
		M3	2.80	0.48 0.44
Mxd	Mya	M4	1.28	0.89
Mxe	My	M5	0.59	1.20
			1.01	1.20
My	Myy	As	1.01	1.00
Mz		spec	1.68	1.00
Mr			1.00	1.00
Vx/Vxa		V1	3.47	0.66
		V2	3.35	0.66
Vxd	Vya	V3	2.74	0.82
Vxe	Vy	V4	1.17	1.20
			1.00	1.20
Vy	Vyy	As	0.34	1.00
Vz		spec	1.70	1.00
Vr			1.00	1.00





cadence



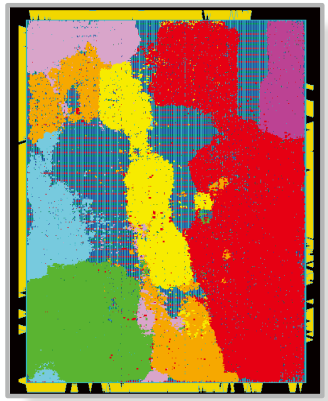
10nm Design Flow Exploration for High-End ARM CPU Implementation

cadence

Phase 1: Process Scaling

From transistor to IP level

- Using a real design spun off from ARM® Cortex®-A73 CPU core
  - Macro-less timing critical dside sub-unit
- Over 20,000 flipflops and 200k instances
- Big enough to present real design challenges
  - Small enough for rapid turnaround time
- Contains uArchitectural critical paths
- Enables experimentation and flow proving
  - Power grid design for primary and switched domains
  - Utilization and routability
  - Standard cell library usage and analysis
  - Full Cadence® platform interaction



cadence

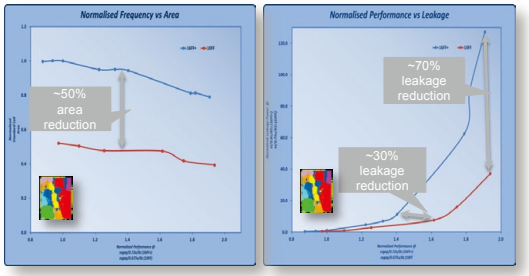
From FoM (Figure-of-Merit) to CPU

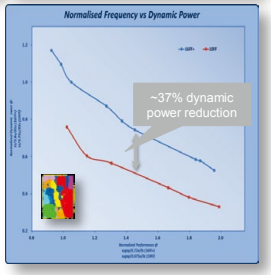
- Area
  - Area is compared using fully routed block
  - Includes complete power grid and power switch network
  - DRC clean from router view
  - Final utilization >70%

  Area reduction of ~50%
- Leakage
  - Compares multiple implementations at each Vt and channel length
  - Demonstrates potential leakage scaling using representative CPU block

  Leakage reduction is 30-70% depending on dominant Vt
- Dynamic power
  - Similar comparisons made for dynamic power
  - Demonstrates potential dynamic scaling based on statistical activity

  Dynamic power reduction of ~37% based on statistical analysis



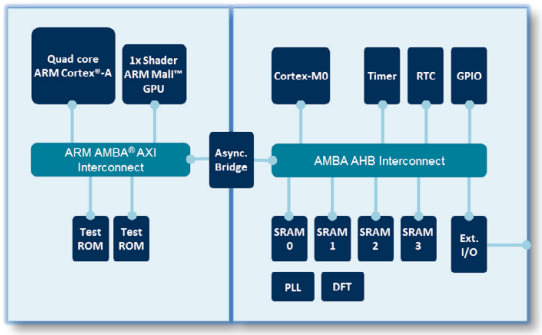


cadence

This slide contains forward-looking statements regarding Cadence's business or products. Actual results may differ materially from the information presented here.

Phase 2: Putting it All Together

From IP to SoC level



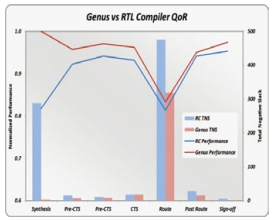
- Taped out to TSMC in December 2015, silicon released and *proven* in the middle of 2016
- Cortex-A73/Mali-T860 signed off with production margins, full manufacturing test support, and power management features

- "Simple" silicon qualification test chip structure
  - Enables PPA benchmark testing on 10nm silicon
  - Manageable test chip complexity
- Real application CPU configuration
  - Quad core latest flagship ARM® Cortex®-A73 cluster
  - Power management
  - Production DFT structures
- Simplified GPU configuration
  - Single Mali™-T860 Shader Core and top level
  - Demonstrates achievable PPA for larger configurations

cadence

Synthesis Strategy

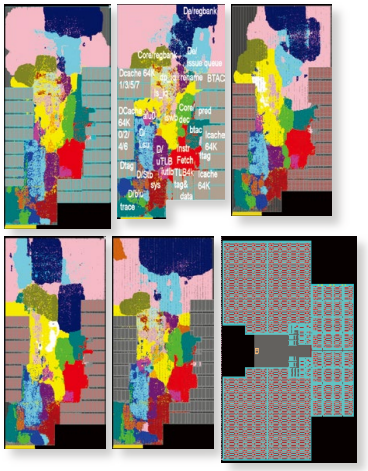
- Exploration to understand power and performance envelope
  - Selecting the right Vt and channel length combination based on performance and power targets
- Analyze library usage for potential bottlenecks
  - Dominant cell type on critical paths
- Tuning of synthesis tool parameters for optimal PPA tradeoffs
- Cadence Genus™ Synthesis Solution advantage
  - Physical aware mapping - Improved correlation and predictability through flow
  - Significant run time and overall QoR benefits - Spatial mode gains optimal tradeoff in between
  - Cortex-A73 CPU synthesis run time over 2X faster (~6 hours) than our previous tool



cadence

Floorplan

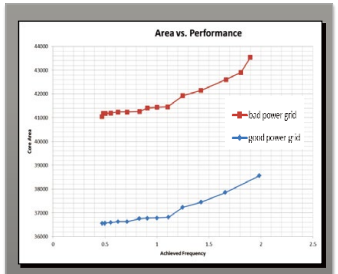
- New process and processor combination
  - Floorplan trials are critical
  - Different SRAM cuts
- Use of placement regions based on uArchitectural feedback
- Macro placement needs to follow FinFET grid
- Must align double pattern layer M2 memory pins to ensure correct coloring



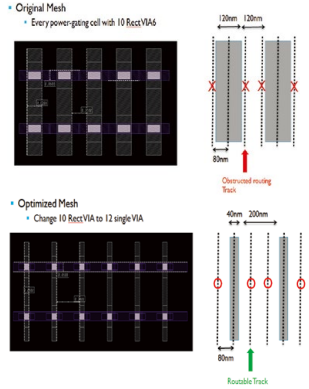
cadence

Power Grid Construction and Refinement

- Incorrect power grid construction will *significantly* impact routing density and performance



Both runs: Cortex-A9 CPU with the same power grid pitches
- Good design has optimally placed M3 straps
- Bad design has suboptimal placement of M3 straps, blocking unnecessary routing tracks



- Reduce R2R WNS 15ps @ postroute stage

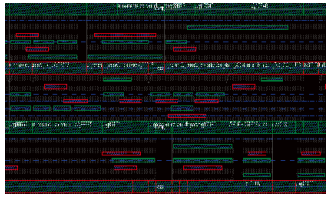
Powerplan type	Placement	CTS	PostCTS	Postroute
Original	62ps	56ps	51ps	47ps
Propose	40ps	40ps	36ps	32ps
Delta	-22ps	-16ps	-15ps	-15ps

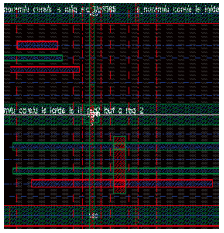
- Less impact on voltage drop
- PASS design rule check

cadence

10nm FinFET Challenges for Power Grid Design

- TSMC 10nm uses horizontal M1 metal direction (HVH), different from 16FF+
  - Any M1 jog or non-preferred orientation strictly prohibited
  - Absence of horizontal M2 standard cell power connection
  - Potential EM and dynamic IR implications
- Traditional “striped” PG methodology used
- Double pattern M1, M2, and M3
- Non-uniform track libraries
  - Retain same rails for VDD and VSS upon symmetrical tracks allocated into every row
- Pitch misalignment: M2 track offset preferred
- Complex DRCs



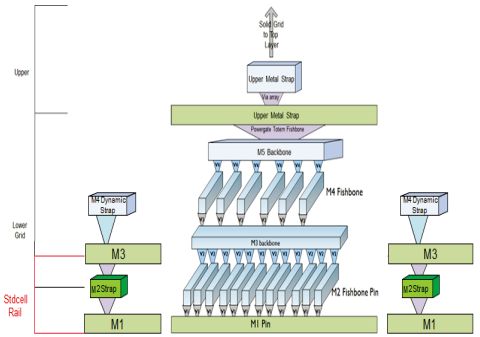


Straightforward usage of P&R tool power grid commands will not create optimal grids

cadence

Power Grid Automation Philosophy

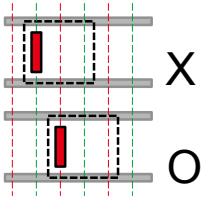
- M3 is added as horizontal backbone to connect M2 fish bones through single V2
  - Concurrently minimize the use of routing tracks and alleviate static IR drop
- V fishbone along with H backbone patterns are stacked to upper metal strap



cadence

N10 Place and Route

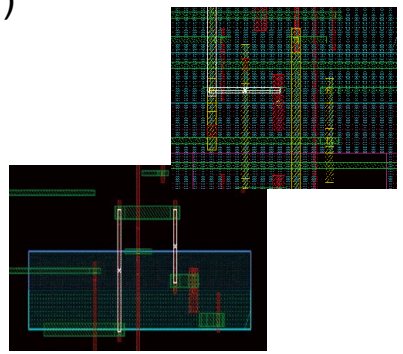
- Place and route uses single Vt/Channel length library
  - Except hold fixing and leakage reclaim
- Row end, top/bottom, and corner cap boundary cells are mandatory
  - Even poly pitch is strictly required for each row
- Timing and power is optimized against dominant corners
  - Balance of run time vs. overall QoR
  - Co-optimization
- Placement is DRC aware and color driven
  - Respects interaction between standard cells
- Control of wire dominated paths is crucial
  - Using placement bounds to constrain potential long paths
  - Layer promotion for long nets to less resistive metal layers
  - Timing-driven global routing is vital to relieve post-route timing drop-off intensified in 10nm

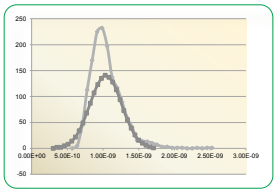


cadence

N10 Place and Route (cont.)

- Stacked via structure for clock NDR routing
  - Clock trunks in M7/M8 and clock leafs in M4-M6
  - Uses default (min) wire width on M2 but ~3X on M3 to avoid signal EM issues
  - Limits M2/M3 usage for short wire pin access only to reduce high-R timing impact
- EM-aware clock tree synthesis to avoid DC EM and reduce IR drop
- Waveform propagation models and AOCV are essential at 10nm
  - Customized AOCV to reduce pessimistic margin

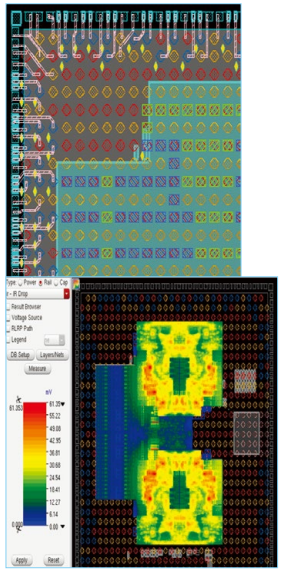




cadence

Signoff

- Full production quality signoff approach
  - TSMC recommended timing margins
  - Signed-off for RC extraction, timing, and power integrity
- Signed off and optimized across multiple PVT and extraction corners
- Cadence Quantus™ QRC Extraction Solution fully supports concurrent multi-corner and explicit color-aware extraction with compelling capacity
- Leakage recovery performed using Cadence Tempus™ Timing Signoff Solution's ECO functionality
- PBA and stage-based OCV signoff critical to avoid over-fixing
- Cadence Voltus™ IC Power Integrity Solution used for all power integrity checks
  - Static/dynamic IR, electro-migration—Early and often to ensure PG structure is correct
  - In-rush current
  - Signoff power analysis

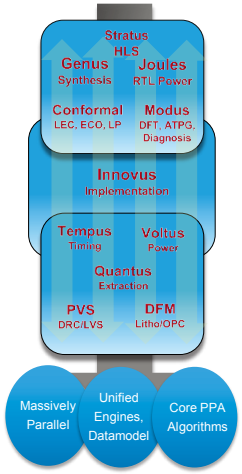


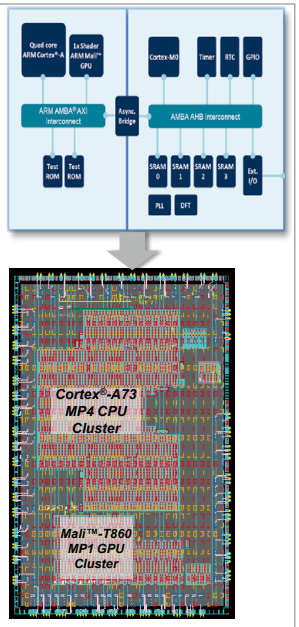
cadence

Design Flow

- Implemented using Cadence tools
- Full production-quality DFT solution
  - Full scan compression
  - Memory BIST
  - At-speed scan and memory BIST support
- Power domain per CPU core
- Full production margins for timing, power, and physical signoff

Cadence Full-Flow Digital Solution



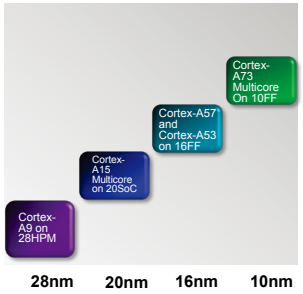


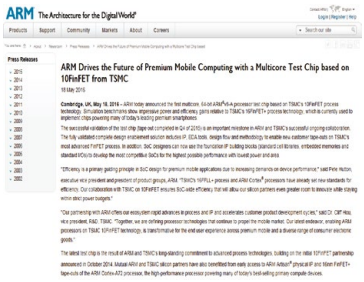
*Cadence Joules™ RTL Power Solution
Cadence Conformal® Technologies
Cadence Modus™ Test Solution
Cadence Innovus™ Implementation System

cadence

Summary and Further Work

- Our proven successful collaboration model accelerates adoption and time-to-market for our partners
- 10nm has arrived and the ecosystem is ready for our next ARM-based chip
- 10nm complexity is *evolutional* from 16nm, some differences but not a step change
- Demonstrating significant power and area advantages
  - Power and area scaling from 16FF+ is in line with expectations
  - Further tool enhancement, flow tweaking, process maturity, and physical IP optimization may further fulfill PPA uplift




ARM The Architecture for the Digital World®

Press Releases

ARM Drives the Future of Premium Mobile Computing with a Multicore Test Chip based on 10FinFET from TSMC

16 May 2016

Cambridge, UK, May 16, 2016 – ARM® today announced the first multicore, 64-bit ARMv8-A processor test chip based on TSMC's 10FinFET process technology. Simulation benchmarks show impressive power and efficiency gains relative to TSMC's 16FinFET+ process technology, which is currently used to implement chips powering many of today's leading premium smartphones.

The successful validation of the test chip (tape-out completed in Q4 of 2015) is an important milestone in ARM and TSMC's successful ongoing collaboration. The fully validated complete design environment solution includes IP, EDA tools, design flow and methodology, to enable new customer tape-outs on TSMC's most advanced FinFET process. In addition, SoC designers can now use the foundation IP building blocks (standard cell libraries, embedded memories and standard I/Os) to develop the most competitive SoCs for the highest possible performance with lowest power and area.

"Efficiency is a primary guiding principle in SoC design for premium mobile applications due to increasing demands on device performance," said Pete Hutton, executive vice president and president of product groups, ARM. "TSMC's 10FFLL+ process and ARM Cortex® processors have already set new standards for efficiency. Our collaboration with TSMC on 10FinFET ensures SoC-wide efficiency that will allow our silicon partners even greater room to innovate while staying within strict power budgets."

"Our partnership with ARM offers our ecosystem rapid advances in process and IP and accelerates customer product development cycles," said Dr. Cliff Hou, vice president, R&D, TSMC. "Together, we are delivering processor technologies that continue to propel the mobile market. Our latest endeavor, involving ARM processors on TSMC 10FinFET technology, is transformative for the end user experience across premium mobile and a diverse range of consumer electronic goods."

The latest test chip is the result of ARM and TSMC's long-standing commitment to advanced process technologies, building on the initial 10FinFET partnership announced in October 2014. Mutual ARM and TSMC silicon partners have also benefited from early access to ARM Artisan® physical IP and 16nm FinFET+ tape-outs of the ARM Cortex-A72 processor, the high-performance processor powering many of today's best-selling premium compute devices.

cadence®

cadence®